BUROS
CENTER FOR TESTING

Conners 4th Edition

Purpose
Designed to “measure symptoms of and impairments associated with Attention-Deficit/Hyperactivity Disorder (ADHD), as well as common co-occurring problems and disorders in youth.”
Population
Ages 6–18.
Publication Dates
1989–2022.
Acronym
Conners 4.
Administration
Individual or group.
Forms, 9
Conners 4 Full (Parent, Teacher, Self-Report); Conners 4 Short (Parent, Teacher, Self-Report); Conners 4 ADHD Index (Parent, Teacher, Self-Report).
Price Data, 2025
$185 per print manual (2022, 448 pages), $5.75 per online use and scoring.
Foreign Language Editions
Parent and Self-Report forms available in U.S. Spanish and French Canadian; Teacher forms available in French Canadian.
Comments
May be administered online or via paper and pencil; scored online.
Authors
C. Keith Conners.
Publisher
Multi-Health Systems, Inc.
a) CONNERS 4 FULL

Population
Ages 6–18 for Parent and Teacher forms; ages 8–18 for Self-Report form.
Time
10–25 minutes for administration.
Scores
2 Validity Indices: Negative Impression Index, Inconsistency Index; 3 Critical and Indicator Items: Severe Conduct, Self-Harm, Sleep Problems Indicator; 6 Content

Scales: Inattention/Executive Dysfunction, Hyperactivity, Impulsivity, Emotional Dysregulation, Depressed Mood, Anxious Thoughts; 3 Impairment and Functional Outcome Scales: Schoolwork, Peer Interactions, Family Life (Parent and Self-Report only); 5 DSM-5 ADHD Symptoms Scales: ADHD Inattentive Symptoms, ADHD Hyperactive/Impulsive Symptoms, Total ADHD Symptoms, Oppositional Defiant Disorder Symptoms, Conduct Disorder Symptoms; ADHD Index.

Comments

Self-Harm Critical Items and/or DSM-5 Conduct Disorder Symptoms scale items can be disabled during administration.

b) CONNERS 4 SHORT

Population

Ages 6–18 for Parent and Teacher forms;
ages 8–18 for Self-Report form.

Time

5–10 minutes for administration.

Scores

Negative Impression Index (validity indicator); 4 Content Scales: Inattention/Executive Dysfunction, Hyperactivity, Impulsivity, Emotional Dysregulation; 3 Impairment and Functional Outcome Scales: Schoolwork, Peer Interactions, Family Life (Parent and Self-Report only); ADHD Index.

c) CONNERS 4 ADHD Index

Population

Ages 6–18 for Parent and Teacher forms; ages 8–18 for Self-Report form.

Time

1–3 minutes for administration.

Score

ADHD Index.

Cross References

For reviews by Sharon Arffa and Thomas M. Dunn of the 3rd Edition, see 18:35; for reviews by Allen K. Hess and Howard M. Knoff of the Revised Edition, see 14:98; see also T5:681 (99 references) and T4:636 (50 references); for reviews by Brian K. Martens and Judy Oehler-Stinnett of the original edition, see 11:87 (83 references).

Review of the Conners 4th Edition by GARY L. CANIVEZ, Canivez EdPsych Consulting; Professor of Psychology, Department of Psychology, Eastern Illinois University, Charleston, IL:

DESCRIPTION

The Conners 4th Edition (Conners 4) is a multi-informant assessment designed to measure symptoms, characteristics, and impairments related to attention-deficit/hyperactivity disorder (ADHD) in children and adolescents ages 6 to 18 years. It is a revision of the Conners 3rd Edition (Conners 3; Conners, 2008) in which some content was retained and modified and new content was added. The Conners 4 has forms completed by parents or guardians (reflecting behaviors and symptoms in the home and community), teachers (reflecting behaviors and symptoms in the school environment), and youth self-report (reflecting child or adolescent perspectives regarding their own behaviors, thoughts, and feelings). Three versions are available that differ in length depending on assessment needs. The most detailed version is the full-length Conners 4 (109–118 items), useful in comprehensive assessment. When the full-length version is impractical or not possible to use, the Conners 4th Edition Short (Conners 4–Short [49–53 items]) is available. Finally, the Conners 4th Edition ADHD Index (Conners 4–ADHD Index [12 items]) is available as a brief and focused assessment that may be used to determine whether further assessment is necessary. The parent and youth self-report forms are available in English, U.S. Spanish, and French-Canadian; the teacher forms are available only in English and French-Canadian. The measure has an extensive and detailed test manual.

The full-length Conners 4 includes numerous scales for various purposes. Response Style indicators help clinicians determine validity of resulting ratings and include the Negative Impression Index, Inconsistency Index, Omitted Items, and Pace (online administration only). The Conners 4–Short includes all but the Inconsistency Index, and the Conners 4–ADHD Index contains none. The full-length Conners 4 includes critical items related to Severe Conduct and Self-Harm and indictor items related to Sleep Problems (not available in the Conners 4–Short or ADHD Index). Content Scales relate to ADHD clinical constructs and include Inattention/Executive Dysfunction, Hyperactivity, Impulsivity, and Emotional Dysregulation; Depressed Mood and Anxious Thoughts are included to measure possible comorbid internalizing problems. All but Depressed Mood and Anxious Thoughts are included in the Conners 4–Short. Impairment and Functional Outcome Scales are included to assess aspects

of clinically significant impact related to Schoolwork, Peer Interactions, and Family Life (parent and self-report only) domains in both the full-length Conners 4 and the Conners 4–Short. DSM-5 Symptom Scales are included in the full-length Conners 4 to assess specific symptom criteria and include ADHD Inattentive Symptoms, ADHD Hyperactive/Impulsive Symptoms, Total ADHD Symptoms, Oppositional Defiant Disorder Symptoms, and Conduct Disorder Symptoms. The Conners 4–ADHD Index is a probability-based score based on machine learning and diagnostic accuracy to differentiate youth with ADHD clinical diagnosis from the general population. Finally, three open-ended questions relating to Impact of Symptoms in Functional Domains, Other Concerns, and Strengths and Skills allow raters to provide additional details following completion of the full-length Conners 4 or the Conners 4–Short. The test manual contains detailed descriptions and a table outlining changes to content between the Conners 3 and the Conners 4.

Two methods of administration are available: a paper-and-pencil administration where the rater completes examiner-printed rating forms from the test publisher's online platform or an online administration (computer, tablet, phone) where the rater completes the electronic form either in the examiner's office or remotely via an emailed link. For paper-and-pencil administration the examiner manually transfers ratings into the online portal where scoring and score reports are completed. Hand scoring is not possible. Conners 4 content was modified to address cultural sensitivity and to include gender-neutral terminology. Rating form instructions and item content were created for a minimum readability of Grade 5 according to the Flesch-Kincaid criteria. For those unable to read because of vision problems or insufficient reading skill, the examiner may read instructions and items as an accommodation. Examiners may disable items related to harm to self or others that might be considered highly sensitive, and in such cases, these items would not be considered "omitted" in related scales nor would they affect their scoring.

Score reports with varying options are available, and scoring options are well described. The default option compares the youth's raw scores to the combined-gender normative sample, a preferred comparison group to retain well-known differential ADHD base rates (and other externalizing and internalizing disorders). Options for comparisons to gender-specific norms are also available. Raw scores also may be compared to the ADHD reference sample (youth with confirmed ADHD diagnosis) in either combined-gender or gender-specific (male, female) groups. Standard scores are T scores (M = 50, SD = 10), and 90% and 95% confidence intervals are available. Confidence intervals (CIs) are obtained score CIs (Glutting et al., 1987), not estimated true score CIs frequently used in tests of intelligence, as described in the test manual and illustrated in case study examples showing symmetrical confidence intervals around the obtained T score. Such CIs are appropriate when the assessment

question relates to the true score at the time of the evaluation. Percentile ranks and T score classifications (low, average, slightly elevated, elevated, very elevated) are also provided. For the ADHD Index, a probability percentage score ranging from 1% to 99% and the likelihood that the individual is more like youth with ADHD (higher percentage) or more like youth in the general population (lower percentage). Descriptive qualitative terms (very low, low, borderline, high, very high) are used to describe probability score ranges. Finally, within-profile comparisons (ipsative scores) provide deviations of the individual's score from the mean across constituent scales, providing a deviation and its statistical significance ($p < .05$), like assessments of strengths and weaknesses within scales in tests of intelligence.

DEVELOPMENT

Conners 4 development is well detailed in the test manual beginning with conceptualization, initial planning, and item development that considered previous versions of the Conners' Rating Scales dating to Conners's initial work in the 1960s. ADHD research published since the introduction of the Conners 3 in 2008 was consulted to identify gaps in assessment and supplemented with market research and Conners 3 user feedback to address perceived needs. Detailed revision goals were provided and included, among other elements, meeting standards of cultural sensitivity, fairness, and language inclusiveness; updating normative samples; better alignment of parent, teacher, and youth self-report forms; and modifying and updating item and scale content. Updated invalid response scales were noted (Inconsistency, Negative Impression), which were assessed using diagnostic utility statistics (Kessel & Zimmerman, 1993). The Positive Impression Scale included in the Conners 3 was dropped. New and modified scales are described with references supporting inclusion, and inclusion of scales to assess impairment and functional outcomes address DSM-5 and IDEA requirements. Various experts noted in the test manual Acknowledgments section provided review and feedback regarding item content, and Spanish (U.S.) and French-Canadian versions were produced by a well-known test translation company, cApStAn, that translated test items and instructions.

Pilot research was conducted on new items with samples of typically developing children (n = 518 [parent], n = 490 [teacher], n = 388 [self-report]) and clinical samples (n = 349 [parent], n = 242 [teacher], n = 245 [self-report]). Detailed demographic descriptions were provided for the samples. Classical test theory (CTT) and item response theory (IRT) methods were used to assess adequacy for inclusion, and poorly performing items were either modified or deleted, resulting in 269 parent, 256 teacher, and 290 self-report items.

TECHNICAL

Standardization

Standardization data collection occurred from May 2019 to March 2020, and participants were recruited from all 50 U.S. states and all 10 Canadian provinces. Details were provided in the test manual regarding disqualifying features that would invalidate responses leading to removal from analyses in 15.3% of parent, 11.7% of teacher, and 19.1% of self-report cases. Final normative samples included 1,560 parent (1,416 U.S., 144 Canadian), 1,560 teacher (1,416 U.S., 144 Canadian), and 1,100 self-report (960 U.S., 140 Canadian) cases (proportional to populations). The samples were stratified on demographic variables of gender, age, race/ethnicity (U.S.: Hispanic, Asian, Black, White, Other; Canada: not visible minority, visible minority), geographic region (U.S.: Northeast, Midwest, South, West; Canada: East, Central, West), and, except for teacher ratings, parental education level (PEL; no high school diploma, high school diploma/GED, some college/associate's degree, bachelor's degree, graduate/professional degree) according to the 2018 U.S. Census American Community Survey and the 2016 Canadian National Census Profile. Gender, not biological birth sex, was specified as male, female, or other (to be specified [i.e., transgender, nonbinary]). Frequencies of each of the stratification variables were crossed with age, and sample percentages were presented with respective Census estimates from the U.S. and Canada; close approximations were observed. The test manual also provided detailed statistics for demographic characteristics of parent raters and teacher raters. The ADHD reference samples included 560 parent, 321 teacher, and 229 self-report ratings and were children and adolescents previously diagnosed with ADHD. ADHD diagnoses included the predominantly inattentive (PI), predominantly hyperactive/impulsive (PH/I), and combined (C) types.

Regression-based continuous norming with statistically smoothed means and standard deviations and square root transformations of positively skewed data (where needed) were provided for both normative and ADHD reference samples. Smoothed means and standard deviations were then used to produce T scores with reportedly no discontinuity among adjacent age groups and less noise from sampling variability. Empirically determined percentiles for the normative sample were produced from the actual score frequency distributions due to non-normality, but theoretical percentiles were produced for the ADHD reference samples due to more normally distributed scores.

Reliability

Four methods for estimating reliability of Conners 4 scores are reported: internal consistency, IRT-based test information function (TIF), interrater agreement, and test-retest stability.

Strong evidence was provided. Internal consistency estimates by age group (2-year intervals) were produced by alpha and omega (model-based estimates based on confirmatory factor analyses [CFA]) coefficients, and item polychoric correlations were used due to the greater accuracy for ordinal item responses. Excellent internal consistency was reported for omega coefficients for all scales across all age and gender groups. Omega coefficients for the normative sample parent ratings ranged from .79 to .98 (Mdn = .94), teacher ratings from .81 to .98 (Mdn = .95), and self-report from .77 to .96 (Mdn = .90). Omega coefficients for the ADHD reference sample parent ratings ranged from .86 to .97 (Mdn = .93), teacher ratings from .81 to .98 (Mdn = .95), and self-report from .71 to .95 (Mdn = .88). Similar ranges and median coefficients were observed for alpha coefficients for all scales across all age and gender groups for the normative and ADHD reference samples. Standard errors of measurement were produced for T scores using omega coefficients.

IRT TIF plots for each scale are presented with parent, teacher, and self-report functions in the same figure for easy comparison. Information and precision appeared high for parent and teacher ratings across all scales and adequate to high for self-ratings. Precision also appeared to range widely from about average performance to 2–3 standard deviations for all scales and raters.

Short-term test-retest stability was investigated for 81 parent raters, 61 teacher raters, and 68 self-reports across retest intervals of 14–30 days. Replication with additional samples is needed. Obtained and range-variation-corrected stability coefficients are presented for all scales, as are means, medians, and standard deviations at Time 1 and Time 2. Corrected stability coefficients across all scales were all statistically significant ($p < .001$) and ranged from .83 to .99 for parents, .81–.97 for teachers, and .63–.86 for self-reports. Small mean differences between Time 1 and Time 2 were observed for parent (−1.6 to 0.3), teacher (−1.0 to 1.0), and self-report (−2.4 to 0.1) ratings. Results showed excellent short-term stability of Conners 4 scales in both pattern and level (McDermott, 1988).

Interrater agreement was assessed by examining raters of the same type (two parents [N = 68], two teachers [N = 34]) or different types (parent-teacher [N = 62], parent-child [N = 62], teacher-child [N = 62]), and obtained correlations and range-variation-corrected correlations are presented with rater means, medians, and standard deviations. Teachers taught different class content, but whether they observed the child in the same classroom at the same time was not specified. For two-parent comparisons, range-corrected correlations across all Conners 4 scales ranged from .58 to .93, and all were statistically significant ($p < .001$), illustrating strong pattern agreement (McDermott, 1988). Mean differences between parent ratings across scales ranged from 0.1 to 1.4 T score points, showing strong level agreement (McDermott, 1988). Range-corrected correlations between teacher ratings across

all Conners 4 scales ranged from .38 to .69, and all were statistically significant ($p < .05$), illustrating generally moderate to strong pattern agreement (McDermott, 1988). Mean differences between teacher ratings across scales ranged from 0.2 to 1.6 T score points, showing strong level agreement (McDermott, 1988). Correlations might have been higher if teachers observed the same child in the same classroom at the same time (see Schaefer et al., 2000). Given the small sample sizes used, replication is needed.

The second interrater agreement study focused on a sample of 62 children/adolescents. Parent, teacher, and self-report ratings were obtained for comparisons between the three different rater pairs. Obtained correlations and range-variation-corrected correlations were presented with rater means, medians, and standard deviations for all Conners 4 scales. Agreement between parent and teacher raters showed corrected correlations ranged from .28 to .79 (Mdn = .46) and mean T score differences ranged from 0.8 to 5.9 points with parent ratings higher than teacher ratings. Agreement between parent and child/adolescent raters showed corrected correlations ranged from .19 to .72 (Mdn = .30) and mean T score differences ranged from 0.2 to 4.6 points with parent ratings higher than child/adolescent ratings. Agreement between teacher raters and child/adolescent raters showed corrected correlations ranged from .11 to .61 (Mdn = .23) and mean T score differences ranged from 0 to 5.0 points with teacher ratings sometimes higher and sometimes lower than child/adolescent ratings. Although generally lower agreement is observed between different rater pairs, this does not necessarily indicate poor “agreement” because a variety of valid reasons for such differences exists such as situation specificity for behavioral expression, opportunity to observe certain behaviors, and varying insights regarding behavior.

Validity

Support for Conners 4 validity was focused on evidence based on internal structure, relations with conceptually related constructs (i.e., comparisons to other tests), and relations with criterion variables (i.e., distinct group differences). Also discussed within validity support was classification accuracy, which might be considered a higher-level requirement than validity given that diagnostic utility requires validity support, but validity support is not sufficient. Historically, many tests show various validity supports but lack diagnostic utility, a feature that is crucial for supporting individual use. Diagnostic utility support is required for evidenced-based assessment methods (Youngstrom 2013, 2014; Youngstrom & Van Meter, 2016).

Support for internal structure of Content Scales was summarized in the Conners 4 manual, but additional details and analyses are necessary. Exclusive use of CFA was reported for

items on the parent, teacher, and self-report forms and 5-, 6-, and 7-factor models were assessed and compared. Commonly used criteria for global model fit (CFI, TLI, RMSEA, SRMR) were reported, and 5-, 6-, and 7-factor models were all well-fitting models for parent and teacher reports for all global fit indexes. For self-report, 5-, 6-, and 7-factor models had adequate global model fit for CFI and TLI, but good model fit for SRMR and RMSEA. Statistical comparisons of 5-, 6-, and 7-factor models showed significant improvement ($p < .01$) from 5- to 6- to 7-factor models for parent, teacher, and self-report, but there were no meaningful differences ($\Delta CFI > .01$) among the three models. The 7-factor model was rejected due to extremely high correlations between Inattention and Executive Dysfunction factors (parent $r = .96$, teacher $r = .98$, self-report $r = .95$) so the 6-factor model was favored. Correlations among the six factors on the parent, teacher, and self-report forms are presented in the test manual and were moderate to high. Not discussed were the correlations between Inattention/Executive Dysfunction, Hyperactivity, Impulsivity, and Emotional Dysregulation, which were higher than their correlations with Depressed Mood and Anxious Thoughts scales in the parent, teacher, and self-report forms, indicating a possible 2-factor higher-order or bifactor representation that might explain these interrelationships. Without such explication and model-based reliability and dimensionality coefficients (see Bornovalova et al., 2020; Reise et al., 2023; Rodriguez et al., 2016a, 2016b; Watkins, 2017), it is not possible to assess the interpretive adequacy of the various Content Scales based on unique portions of variance captured. Standardized coefficients (loadings) for items with associated factors were uniformly high and supported assignments for parent, teacher, and self-report forms.

Convergent (and discriminant) validity related to Conners 4 scales was assessed with comparisons to the Conners 3; the Behavior Assessment System for Children, Third Edition (BASC-3; Reynolds & Kamphaus, 2015); the Comprehensive Executive Function Inventory (CEFI; Naglieri & Goldstein, 2013); and the Weiss Functional Impairment Rating Scales (WFIRS; Weiss et al., 2018). In a study of 85 parents, 58 teachers, and 81 children/adolescents who completed both the Conners 4 and Conners 3 within 1 month, moderate to high and statistically significant ($p < .001$) correlation coefficients (rs ranged .42 to .98) between the same or similar scales were reported for the parent, teacher, and self-report forms. In a study of 71 parents, 120 teachers, and 77 children/adolescents completing the Conners 4 and the BASC-3 within 14 days, statistically significant ($p < .001$) moderate to high correlation coefficients (rs ranged .57 to .96) between the same or similar scales were reported for the parent, teacher, and self-report forms, supporting convergent validity. Completion of the Conners 4 and CEFI by 137 parents, 122 teachers, and 54 children/adolescents within 4 weeks produced statistically significant ($p < .001$) moderate to high correlation coefficients (rs ranged $-.56$ to $-.78$) between similar scales and supported convergent validity. (Higher CEFI scores represent greater functioning.) Comparisons of

Conners 4 Impairment & Functional Outcome Scales and WFIRS domains for parents (N = 558–560) and children/adolescents (N = 140) found statistically significant ($p < .001$) moderate to high correlation coefficients (rs ranged .52 to .86), further supporting convergent validity.

Clinical group differences (distinct group differences) validity was examined for Conners 4 parent, teacher, and self-report forms by comparing scores between youth previously diagnosed with ADHD PI, ADHD PH/I, ADHD (C), depression (major depressive disorder, persistent depressive disorder), generalized anxiety disorder, and disruptive disorders (oppositional defiant disorder, conduct disorder, intermittent explosive disorder) and those in the general population using ANOVAs, Tukey HSD post hoc tests, and various effect size estimates. ADHD clinical groups showed significant differences with large effect sizes compared to the general population for parent, teacher, and self-report ADHD oriented scales. Due to large effect sizes, separation of group score distributions suggests the possibility for meaningful diagnostic utility necessary for individual clinical use.

Test developers recognized that distinct group differences studies or clinical group differences are necessary but not sufficient indicators for clinical utility and used binary logistic regression and diagnostic accuracy conditional probabilities recommended by Kessell and Zimmerman (1993) for these analyses. Various combinations of Conners 4 scales (ADHD based Content Scales + Impairment & Functional Outcome Scales, DSM-5 Total ADHD Symptoms Scale + Impairment & Functional Outcome Scales, DSM-5 ADHD Inattentive Symptoms Scale + Impairment & Functional Outcome Scales, and DSM-5 Hyperactive/Impulsive Symptoms Scale + Impairment & Functional Outcome Scales) were used as logistic regression predictors with demographically matched (age, gender, race/ethnicity, PEL) samples of youth with ADHD diagnoses and youth from the general population (50% base rate). Conditional probabilities and kappa coefficients are fully reported with emphasis on positive predictive value (PPV) and negative predictive value (NPV), indexes for ruling in or ruling out classification, respectively. Diagnostic utility for the parent form appeared better than for the teacher and self-report forms. The test author noted conditional probability estimates PPV and NPV are influenced by base rates (as well as cut scores), so PPV and NPV estimates were calculated for parent, teacher, and self-report forms with varying population base rates of 10%, 60%, 70%, and 80%. With lower base rates, PPV was low while NPV was high, and as base rates increased from 50% to 80%, PPV increased while NPV decreased.

COMMENTARY

The Conners 4 is well designed and normed, and incorporation of electronic administration,

scoring, and analyses is useful. Strong preliminary evidence for score reliability was presented. Convergent and distinct group differences validity was also supportive. Although general support appears in the structural validity of the scales, because factor intercorrelations were moderate to high, further assessment of a hierarchical structure (higher-order) or bifactor representation with resulting model-based reliability and dimensionality estimates are needed to further assess adequacy of scale interpretations. Also, while the inclusion of diagnostic accuracy analyses was welcome and it was acknowledged that base rates affect such estimates, cross-validation is required as binary logistic regression will maximize group discrimination. Assessment with independent samples is needed. Receiver operating characteristic (ROC) curve analyses (Swets, 1996; Treat & Viken, 2012), which are not affected by base rates or cut scores, also should be considered in future studies. It is hoped that the test publisher will grant independent researchers access to data from normative and clinical samples to assist in furthering such research. Such assessments will help evaluate the evidence-based use of the Conners 4. Given relatively small samples for some reliability and validity studies reported in the test manual, replication by independent researchers is recommended.

SUMMARY

The Conners 4 represents a well-designed and easy to use measure of ADHD and related symptoms that has good norms and ample preliminary evidence for score reliability, validity, and utility. It is an excellent revision that should provide users confident assessment of ADHD symptoms for comparison to other features in screening, clinical assessment, and research. As with all tests, results presented in test manuals should be considered preliminary, and replication with independent research will help further assess technical adequacy.

REVIEWER'S REFERENCES

Bornovalova, M. A., Choate, A. M., Fatimah, H., Petersen, K. J., & Wiernik, B. M. (2020). Appropriate use of bifactor analysis in psychopathology research: Appreciating benefits and limitations. Biological Psychiatry, 88(1), 18–27. https://doi.org/10.1016/j.biopsych.2020.01.013

Conners, C. K. (2008). Conners 3rd Edition. Multi-Health Systems.

Glutting, J. J., McDermott, P. A., & Stanley, J. C. (1987). Resolving differences among methods of establishing confidence limits for test scores. Educational and Psychological Measurement, 47(3), 607–614. https://doi.org/10.1177/001316448704700307

Kessel, J. B., & Zimmerman, M. (1993). Reporting errors in studies of the diagnostic performance of self–administered questionnaires: Extent of the problem, recommendations for standardized presentation of results, and implications for the peer review process. Psychological Assessment, 5(4), 395–399. https://doi.org/10.1037/1040-3590.5.4.395

McDermott, P. A. (1988). Agreement among diagnosticians or observers: Its importance and determination. Professional School Psychology, 3(4), 225–240. https://doi.org/10.1037/h0090563

Naglieri, J. A., & Goldstein, S. (2013). Comprehensive Executive Function Inventory. Multi-Health Systems.

Reise, S. P., Mansolf, M., & Haviland, M. G. (2023). Bifactor measurement models. In R. H. Hoyle (Ed.), Handbook of structural equation modeling (2nd ed., pp. 329–348). Guilford Press.

Reynolds, C. R., & Kamphaus, R. W. (2015). Behavior Assessment System for Children, Third Edition. Pearson.

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016a). Applying bifactor statistical indices in the evaluation of psychological measures. Journal of Personality Assessment, 98(3), 223–237. https://doi.org/10.1080/00223891.2015.1089249

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016b). Evaluating bifactor models: Calculating and interpreting statistical indices. Psychological Methods, 21(2), 137–150. https://doi.org/10.1037/met0000045

Schaefer, B. A., Watkins, M. W., & Canivez, G. L. (2000). Cross-context agreement of the Adjustment Scales for Children and Adolescents. Journal of Psychoeducational Assessment, 19(2), 123–136.

Swets, J. A. (1996). Signal detection theory and ROC analysis in psychological diagnostics: Collected papers. Erlbaum.

Treat, T. A., & Viken, R. J. (2012). Measuring test performance with signal detection theory techniques. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), APA handbook of research methods in psychology: Foundations, planning, measures, and psychometrics (Vol. 1, pp. 723–744). American Psychological Association. https://doi.org/10.1037/13619-038

Watkins, M. W. (2017). The reliability of multidimensional neuropsychological measures: From alpha to omega. The Clinical Neuropsychologist, 31(6–7), 1113–1126. https://doi.org/10.1080/13854046.2017.1317364

Weiss, M. D., McBride, N. M., Craig, S., & Jensen, P. (2018). Conceptual review of measuring functional impairment: Findings from the Weiss Functional Impairment Rating Scale. Evidence-Based Mental Health, 21(4), 155–164. https://doi.org/10.1136/ebmental-2018-300025

Youngstrom, E. A. (2013). Future directions in psychological assessment: Combining evidence-based medicine innovations with psychology's historical strengths to enhance utility. Journal of Clinical Child & Adolescent Psychology, 42(1), 139–159. https://doi.org/10.1080/15374416.2012.736358

Youngstrom, E. A. (2014). A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: We are ready to ROC. Journal of Pediatric Psychology, 39(2), 204–221. https://doi.org/10.1093/jpepsy/jst062

Youngstrom, E. A., & Van Meter, A. (2016). Empirically supported assessment of children and adolescents. Clinical Psychology: Science and Practice, 23(4), 327–347. https://doi.org/10.1037/h0101738

Cite this review

Canivez, G. L. (in press). [Test review of Conners 4th Edition]. In J. F. Carlson, K. F. Geisinger, J. L. Jonson, & N. A. Anderson (Eds.), The twenty-second mental measurements yearbook. Retrieved from https://marketplace.unl.edu/buros/